

The guilty knowledge test and the modified stroop task in detection of deception: An explorative study

Citation for published version (APA):

Engelhard, I. M., Merckelbach, H. L. G. J., & van den Hout, M. A. (2003). The guilty knowledge test and the modified stroop task in detection of deception: An explorative study. *Psychological Reports*, 92(2), 683-691. <https://doi.org/10.2466/pr0.2003.92.2.683>

Document status and date:

Published: 01/01/2003

DOI:

[10.2466/pr0.2003.92.2.683](https://doi.org/10.2466/pr0.2003.92.2.683)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 04 Dec. 2019

THE GUILTY KNOWLEDGE TEST AND THE MODIFIED
STROOP TASK IN DETECTION OF DECEPTION:
AN EXPLORATORY STUDY¹

IRIS M. ENGELHARD

HARALD MERCKELBACH

*Department of Medical, Clinical and
Experimental Psychology*

Department of Psychology and Faculty of Law

Maastricht University

MARCEL A. VAN DEN HOUT

*Department of Medical, Clinical, and Experimental Psychology
Maastricht University*

Summary.—The current study tested whether a simple Stroop paradigm can be used to detect deceptive behavior. 40 university students (34 women), half of whom committed a mock crime, were administered a Guilty Knowledge Test and modified Stroop task to detect guilt or innocence. The Guilty Knowledge Test is a well-known psychophysiological detection method, which consists of multiple-choice questions about details of the crime while skin conductance is recorded. Subjects possessing guilty knowledge are expected to show enhanced differential responses to the relevant stimuli. The modified Stroop task required color-naming of colored words related to the mock crime or an irrelevant crime. Each version of the Stroop task was presented in story form. Subjects possessing guilty knowledge were expected to produce larger reaction times to the relevant version relative to the irrelevant version. The test correctly identified 100% of innocent participants and 78% of guilty participants. In contrast, Stroop interference, i.e., reaction times for irrelevant crime details subtracted from those for mock crime details, did not differentiate between the two groups, suggesting that the story form of the Stroop paradigm is not suitable for lie detection.

One of the best researched lie detection methods is the Guilty Knowledge Test (Lykken, 1959, 1981/1998; Ben-Shakhar, Bar-Hillel, & Leiblich, 1986; Iacono & Patrick, 1988; Honts & Quick, 1995). Using psychophysiological recordings, this test evaluates the presence of knowledge that would only be accessible to a guilty person and the police. It consists of multiple-choice questions about crime-details that are each followed by a series of items: the "guilty knowledge" item combined with incorrect alternatives. Subjects are typically instructed to respond "no" to each item (e.g., Bradley & Rettinger, 1992). All items are intended to be equally plausible to the innocent. A consistently stronger physiological response to the correct item, relative to the incorrect alternative items, is thought to be diagnostic of

¹Please address correspondence to Iris M. Engelhard, Department of Medical, Clinical, and Experimental Psychology, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands or e-mail (i.engelhard@dep.unimaas.nl).

guilty knowledge. The physiological response typically recorded is skin conductance, which is a sensitive measure of autonomic arousal (e.g., Bradley, & Janisse, 1981), and often this measure is combined with other autonomic measures such as respiration line length (Elaad, Ginton, & Jungman, 1992). Laboratory research shows that the Guilty Knowledge Test is highly accurate: hit rates circle around 90% for guilty and innocent participants (for reviews, see Iacono & Patrick, 1988; Ben-Shakhar & Furedy, 1990). Field studies (Elaad, 1990; Elaad, *et al.*, 1992) also showed high rates for innocent individuals (94–98%) but showed lower hit rates for guilty individuals (42–76%). As acknowledged by the authors, this may be due to the limited number of questions that were used. It may also reflect one limitation of the Guilty Knowledge Test in real life: some of the guilty may be unaware of pertinent crime details. The test relies on guilty individuals' perception and memory of crime details, and these may be restricted by various factors such as emotional state, complexity of the event, postevent elaboration, and time elapsed since the event (Schacter, 1996). Yet, for innocent subjects, the test provides a statistical precaution against misclassification, but only if critical information has not been leaked out. The more "good" questions included, the lower the probability that an innocent person fails (Lykken, 1988). A question is "good" if all alternatives are equally plausible to an innocent person and a guilty person has at least an 80% chance of hitting on that same item (Lykken, 1981/1998).

A more practical limitation of the Guilty Knowledge Test is that it depends on psychophysiological techniques and expertise. Thus, a simple lie detection tool that does not require sophisticated technology would be of great significance. Moreover, this might add more information, so that information from various sources could be aggregated to produce a more efficient detection measure. One interesting candidate in this respect is the modified Stroop paradigm, which taps preferential processing of word meanings (Williams, Mathews, & MacLeod, 1996). In this test, neutral and threatening words are presented that are printed in different colors. Subjects are instructed to name the colors of these words as quickly as possible and to ignore the meaning of the words. Since attentional resources are limited, color-naming slows down when the meaning of the words demands attention. Usually an interference score is calculated by subtracting the reaction time for neutral words from that for threatening words. In general, individuals barely show interference (e.g., Mathews & MacLeod, 1985). Anxious patients, on the other hand, show selective interference, i.e., they are slower in color-naming threatening words than neutral words (e.g., McNally, Kaspi, Riemann, & Zeitlin, 1990). Such interference effects are especially robust when 'threatening' words are specifically related to issues that concern the individual. This might also apply to guilty individuals under interrogation. If verbal cues are

presented that pertain to the pertinent crime, guilty persons would be expected to attend selectively to those cues. Perhaps, then, the Stroop paradigm may give a subtle measure of (extra) processing resources allocated to guilty knowledge cues. Interestingly, it has been suggested that the Stroop paradigm may be used in procedures involving clinical selection. For example, McNally, *et al.* (1990, p. 402) pointed out that it "may be capable of distinguishing between genuine and factitious PTSD [posttraumatic stress disorder]." Note that the Stroop paradigm is a relatively simple and straightforward procedure in that it only involves presentation of verbal cues and measurements of reaction times, i.e., color-naming latencies. Admittedly, one could argue that reaction time performance is sensitive to strategic control, but recent studies (e.g., Seymour, Seifert, Shafto, & Mosmann, 2000) show that this potential limitation should not be exaggerated.

In this exploratory study, we tested whether the emotional Stroop paradigm has potential as a lie detection tool. University students, half of whom committed a mock crime, were administered the Guilty Knowledge Test and modified Stroop task to detect guilt or innocence. The Stroop task consisted of a Guilty Knowledge version with details of the mock crime and a Control version about an irrelevant crime. Both Stroop versions were presented in story-form, since crime details are more easily described by sentences depicting complex situations than by single words. We anticipated that "guilty" participants would be occupied with crime details and therefore would be slower in color-naming the Guilty Knowledge version than the Control version. We did not expect innocent participants to show this selective interference. Thus, the aim of this study was to explore (1) whether Stroop interference can differentiate innocent and guilty participants in a mock crime and (2) if so, whether the Stroop test performs as well as the Guilty Knowledge Test.

METHOD

Participants

Forty undergraduate students (34 women), with a mean age of 19.8 yr. (range 18 to 24), participated for a reward of \$15. Half of the participants were randomly assigned to a "guilty" group and half to an "innocent" group, with men and women equally represented in both groups. None had taken psychology or psychophysiology classes.

Apparatus

The physiological data were recorded with a Picker Schwarzer ED 14 at a chart speed of 10 mm/sec. Using the method of constant voltage (.5 V), skin conductance, i.e., electrodermal, responses expressed in microsiemens were measured with three Beckman Ag-AgCl electrodes (5 mm diameter).

Electrodes were filled with isotonic paste and connected to a Picker Schwarzer EDA 48. Before attachment, participants rinsed their hands with warm water, and the skin was cleaned with a cotton pad dipped in distilled water. Electrodes were then attached to the medial phalanges of the first and second fingers and the medial top of the nondominant hand. Sensitivity was adjusted individually. To control for electrodermal artefacts due to respiratory irregularities, respiration was measured with a Picker Schwarzer respiration belt attached around the chest. An Olivetti M250 microcomputer with a color monitor presented the Stroop and recorded response latencies per version in msec.

Procedure

A research assistant gave each participant written instructions, which were then carried out. The guilty group simulated a realistic theft (based on the procedure described by Raskin, 1983). They entered an office, looked for a wallet in a desk, took money from it (50 Dutch guilders), left the office quickly, and returned to the assistant, who retrieved the money. They were instructed not to tell anyone what they had done and to make up an alibi if they got caught. Innocent participants were merely told that a crime had been committed and that they were among the suspects. They waited for further instructions. Next, both groups were told that they were suspects in a crime and that an examiner would administer two lie detection tests to detect guilt or innocence. They were told "The examiner doesn't know if you're innocent or guilty. Don't confess and don't disclose any details of the crime. Be cooperative so that you won't provoke suspicion. If you are guilty, try to lie effectively during the lie detection so that you'll be declared innocent. You'll get a bonus of 15 guilders if you are declared innocent by the examiner." Participants were introduced to the examiner, who was blind to guilt or innocent status. The order of the Guilty Knowledge Test and the Stroop was counterbalanced within each group, so that half of each group started with the Guilty Knowledge Test followed by the Stroop and the other half had the reversed order.

Guilty Knowledge Test

The examiner attached electrodes and respiration belt and gave a brief explanation about the polygraph. In addition, participants were told "A lie detector can, so to speak, look into your soul. If you're guilty, it will show your responses and identify you as guilty. There's nothing that you can do about that. If you're innocent, you don't need to worry. You won't have any substantial responses and the lie detector will identify you as innocent." There were seven questions, each followed by a sequence of six possible answer options, for example, "You stole money from the wallet. Was it: twenty guilders? one hundred guilders? ten guilders? fifty guilders? thirty guilder-

ders? seventy guilders?" Correct stimuli were positioned near the middle. Questions and stimuli were reviewed with participants and they were requested to attend to each stimulus and give a standard response ("no") to all stimuli. Then the seven questions and their options were presented twice. A new option was presented when the electrodermal response to the previous option had returned to baseline level. Polygraph registrations of the first series were inspected to identify guilt or innocence. The first option of each question served as buffer and was excluded from scoring. For each option, we calculated the increase in electrodermal response occurring within 5 sec. after the participant's verbal response. The magnitude of the electrodermal response to the correct item was compared to that of the responses to the other four items of the set. It was assigned 2 if it had the largest amplitude of that set, 1 if it had the second largest magnitude, or 0 for any other response magnitude (see Lykken, 1959). The total possible range was 0 to 14, and participants with a score of 7 or higher were classified as guilty; those with a lower score were classified as innocent. Individuals showing reduced electrodermal activity, i.e., responses to ≤ 3 questions, were excluded from the analyses (see also Iacono, Boisvenu, & Fleming, 1984; Iacono, Cerri, Patrick, & Fleming, 1992). For practical reasons, we focused on electrodermal reactivity, as this is a sensitive parameter (Honts, Devitt, Winbush, & Kirchner, 1996).

Stroop Task

The modified Stroop consisted of a Guilty Knowledge version with details of the mock crime and a Control version with irrelevant crime details. The Guilty Knowledge version referred to the same critical details as the correct answer options of the Guilty Knowledge Test questions. A sample item of the Guilty Knowledge Stroop was "The wallet contained fifty guilders." The Control version concerned the theft of a bike and served to provide a within-subject baseline in color-naming latencies. The two versions were matched for number of words, i.e., eight statements, 52 words total, and Dutch syllables and were presented on a computer screen. The stories appeared in the center of the screen, 2000 msec. after a little dot marker. Words were displayed in capital letters of 4 mm, printed in four different randomly assigned colors, i.e., red, blue, yellow, or green. Participants were instructed to name the colors as quickly as possible, from left to right, top to bottom, and to ignore the meaning of the words. They were told to correct themselves as quickly as possible if they misnamed a color. After a practice trial, each entire Stroop was presented twice (the order was Control, Guilty Knowledge, Control, Guilty Knowledge) to cancel out possible practice and fatigue effects. The computer recorded total reaction times for each version in milliseconds. Prior to the color-naming, participants read each

version (printed in black and white on paper) to enhance semantic activation.

RESULTS

As there was evidence for carryover effects between the two paradigms,² only results of the between-subjects component of the experiment were analyzed. This included each paradigm presented in the first position, with 20 participants (11 guilty) having the Guilty Knowledge Test and 20 participants (9 guilty) having the Stroop task.

Guilty Knowledge Test

All polygraph charts were blindly rescored after the experiment. Three participants were excluded from the Guilty Knowledge Test because their electrodermal activity was too minimal. Since the Guilty Knowledge Test was administered twice, accuracy was computed per series. For the first series, accuracy was 100% (8/8) for the innocent and 78% (7/9) for the guilty. For the second series, accuracy was 88% for the innocent and 63% for the guilty.

Stroop Task

Errors in color naming were rare and guilty and innocent participants did not differ in this respect. Hence, errors are not considered further. Since both Stroop versions were administered twice, mean reaction times (milliseconds) were averages for each version and an interference score was calculated by subtracting mean reaction times. Contrary to our expectations, there were no significant differences in interference scores between innocent and guilty individuals. That is, relative to the innocent group ($M=1186$, $SD=1764$), the guilty group ($M=391$, $SD=1581$) did not show more interference when color-naming latencies for the Control Stroop were subtracted from those for the Guilty Knowledge Stroop ($F_{1,18}=1.10$, ns, one-tailed test). Although the guilty group was somewhat slower in color-naming the Guilty Knowledge version compared to the innocent group ($M=34862$, $SD=3438$ versus $M=31470$, $SD=5075$), this difference did not attain significance ($F_{1,18}=2.91$, ns, one-tailed test). The guilty group also exhibited somewhat slower latencies for the Control version than did the innocent group ($M=34471$, $SD=3526$ versus $M=30284$, $SD=5650$), and this difference was statistically significant ($F_{1,18}=3.73$, $p=.03$, one-tailed test).

²For example, three innocent participants who were administered the Guilty Knowledge Test *after* the Stroop were misclassified as guilty. During debriefing, it appeared that they were aware of four to six correct Guilty Knowledge Test items due to the preceding Stroop test. In addition, guilty participants administered the Guilty Knowledge Test *after* the Stroop exhibited stronger electrodermal signs of guilty knowledge compared to guilty participants starting with the Guilty Knowledge Test.

DISCUSSION

The findings of this study can be summarized as follows. To begin with, in line with a number of previous laboratory studies (see review by Lykken, 1981/1998; Honts & Quick, 1995) our results show that the Guilty Knowledge Test accurately classified innocent participants since all of them were correctly identified in the first series. Guilty participants were less accurately classified, with hit rates of the first Guilty Knowledge Test series being 78%. Secondly, and most important, results of the Stroop task indicated that, in its present form, it is not suitable as a lie detector.

The Guilty Knowledge Test data clearly illustrate that guilty knowledge is necessary for successful detection. However, guilty knowledge does not appear sufficient, as was illustrated by three findings. First, two guilty participants were misclassified even though they correctly recalled all Guilty Knowledge Test items during the debriefing. When asked if they had done anything to beat the test, one participant could not give an explanation and simply claimed she was innocent because she had merely followed instructions, and the other reported that he relaxed during the test and tried to ignore the content of the stimuli. Second, two guilty participants rapidly habituated and were declared inconclusive. They reported that halfway during the test, they began to imagine additional items for each question and directed their attention to these particular items and responded to the real items in an automatic way. This is consistent with a study about mental countermeasures, which showed that attempts to ignore the content of the Guilty Knowledge Test decrease detection of guilt (Elaad & Ben-Shakhar, 1991; see also Honts, *et al.*, 1996). It is worrying that such countermeasures cannot easily be made visible. Third, due to the carryover effect, innocent participants who started with the Stroop were aware of crime-relevant information during the subsequent Guilty Knowledge Test. Three of them were incorrectly classified as guilty. In accordance with previous findings (Bradley & Rettinger, 1992; Iacono, *et al.*, 1992; Ben-Shakhar, Gronau, & Elaad, 1999), this suggests that at least some of the innocent people may be vulnerable to misclassification when exposed to crime details. Also, hit rates for guilty participants were inflated after priming by the Stroop. Previous research (O'Toole, Yuille, Patrick, & Iacono, 1994) showed that rehearsal of crime-details increases recognition of critical items and guilty knowledge scores.

Whatever the limitations of our results for the Guilty Knowledge Test, they also show that even a suboptimal version of this lie detection method, i.e., exclusive reliance on electrodermal activity, reaches reasonable accuracy rates in classifying guilty and innocent participants. This, then, is a standard against which the performance of the Stroop test can be evaluated. It is clear that our version of the Stroop test did not correctly classify innocent and

guilty participants. In fact, the results of our Stroop paradigm were extremely disappointing and stand in sharp contrast to the results of Seymour, *et al.* (2000) who also employed reaction times as a lie detection parameter. One major difference between their study and ours is that reaction times in our study were based on a small number of observations, i.e., four stories, whereas Seymour, *et al.* relied on a relatively large number of reaction time trials. Our mode of representation introduced a large amount of noise. One could employ a Stroop variant with a large series of individual target words and colors for which separate reaction times could be obtained. However, isolated words will often not convey crime-specific details. Thus, although from a practical point of view, a Stroop paradigm is easier to run than a Guilty Knowledge Test, it is difficult to see how it could acquire similar diagnostic potential.

In summary, while the present study concurs with previous studies showing that the Guilty Knowledge Test may be an effective tool for detecting guilt, it also indicates some serious boundary conditions for the Guilty Knowledge Test, e.g., carryover effects, rapid habituation of electrodermal activity. The Stroop results demonstrated that the story-form is not suitable for use in lie detection. Given the relatively large standard deviations, this measure appears to be too rough. Stroop variants that include reaction times to guilty knowledge presented on a word-by-word basis may be more sensitive. This issue warrants further study. However, an obvious limitation of this variant is that it has little external validity. In real life, it would be extremely difficult to tap complex guilty knowledge with single words.

REFERENCES

- BEN-SHAKHAR, G., BAR-HILLEL, M., & LIEBLICH, I. (1986) Trial by polygraph: scientific and juridical issues in lie detection. *Behavioral Sciences & the Law*, 4, 459-479.
- BEN-SHAKHAR, G., & FUREDY, J. J. (1990) *Theories and applications in the detection of deception: a psychophysiological and international perspective*. New York: Springer-Verlag.
- BEN-SHAKHAR, G., GRONAU, N., & ELAAD, E. (1999) Leakage of relevant information to innocent examinees in the GKT: an attempt to reduce false-positive outcomes by introducing target stimuli. *Journal of Applied Psychology*, 84, 651-660.
- BRADLEY, M. T., & JANISSE, M. P. (1981) Accuracy demonstrations, threat, and the detection of deception: cardiovascular, electrodermal, and pupillary measures. *Psychophysiology*, 18, 307-314.
- BRADLEY, M. T., & RETTINGER, J. (1992) Awareness of crime-relevant information and the Guilty Knowledge Test. *Journal of Applied Psychology*, 77, 55-59.
- ELAAD, E. (1990) Detection of guilty knowledge in real-life criminal investigations. *Journal of Applied Psychology*, 75, 521-529.
- ELAAD, E., & BEN-SHAKHAR, G. (1991) Effects of mental countermeasures on psychophysiological detection in the Guilty Knowledge Test. *International Journal of Psychophysiology*, 11, 99-108.
- ELAAD, E., GINTON, A., & JUNGMAN, N. (1992) Detection measures in real-life criminal guilty knowledge tests. *Journal of Applied Psychology*, 77, 757-767.
- HONTS, C. R., DEVITT, M. K., WINBUSH, M., & KIRCHER, J. C. (1996) Mental and physical countermeasures reduce the accuracy of the concealed knowledge test. *Psychophysiology*, 33, 84-92.

- HONTS, C. R., & QUICK, B. D. (1995) The polygraph in 1995: progress in science and the law. *North Dakota Law Review*, 71, 987-1020.
- IACONO, W. G., BOISVENU, G. A., & FLEMING, J. A. E. (1984) Effects of diazepam and methylphenidate on the electrodermal detection of guilty knowledge. *Journal of Applied Psychology*, 69, 289-299.
- IACONO, W. G., CERRI, A. M., PATRICK, C. J., & FLEMING, J. A. E. (1992) Use of antianxiety drugs as countermeasures in the detection of guilty knowledge. *Journal of Applied Psychology*, 77, 60-64.
- IACONO, W. G., & PATRICK, C. J. (1988) Assessing deception: polygraph techniques. In R. Rogers (Ed.), *Clinical assessment of malingering and deception*. New York: Guilford. Pp. 205-233.
- LYKKEN, D. T. (1959) The GSR in the detection of guilt. *Journal of Applied Psychology*, 43, 385-388.
- LYKKEN, D. T. (1981/1998) *A tremor in the blood: uses and abuses of the lie detector*. New York: McGraw-Hill.
- LYKKEN, D. T. (1988) Detection of guilty knowledge: a comment on Forman and McCauley. *Journal of Applied Psychology*, 73, 303-304.
- MATHEWS, A., & MACLEOD, C. (1985) Selective processing of threat cues in anxiety states. *Behaviour Research and Therapy*, 23, 563-569.
- MCNALLY, R. J., KASPI, S. P., RIEMANN, B. C., & ZEITLIN, S. B. (1990) Selective processing of threat cues in posttraumatic stress disorder. *Journal of Abnormal Psychology*, 99, 398-402.
- O'TOOLE, D., YUILLE, J. C., PATRICK, C. J., & IACONO, W. G. (1994) Alcohol and the physiological detection of deception: arousal and memory influences. *Psychophysiology*, 31, 253-263.
- RASKIN, D. C. (1983) The scientific basis of polygraph techniques and their uses in the judicial process. In A. Trankell (Ed.), *Reconstructing the past: the role of psychologists in criminal trials*. Deventer, The Netherlands: Kluwer. Pp. 317-371.
- SCHACTER, D. L. (1996) *Searching for memory: the brain, the mind and the past*. New York: Basic Books.
- SEYMOUR, T. L., SEIFERT, C. M., SHAFTO, M. G., & MOSMANN, A. L. (2000) Using response time measures to assess "guilty knowledge." *Journal of Applied Psychology*, 85, 30-37.
- WILLIAMS, J. M. G., MATHEWS, A., & MACLEOD, C. (1996) The emotional Stroop task and psychopathology. *Psychological Bulletin*, 120, 3-24.

Accepted February 7, 2003.